Physical Communication 21 (2016) 10–18

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers

Ameen Abdelmutalab, Khaled Assaleh, Mohamed El-Tarhuni *

Electrical Engineering Department, American University of Sharjah, Sharjah, United Arab Emirates

ARTICLE INFO

Article history:
Received 6 November 2015
Received in revised form 17 July 2016
Accepted 24 August 2016
Available online 5 September 2016

Keywords:
Modulation classification
Hierarchical polynomial classifiers
High order cumulants
Adaptive modulation

ABSTRACT

In this paper, a Hierarchical Polynomial (HP) classifier is proposed to automatically classify M-PSK and M-QAM signals in Additive White Gaussian Noise (AWGN) and slow flat fading environments. The system uses higher order cumulants (HOCs) of the received signal to distinguish between the different modulation types. The proposed system divides the overall modulation classification problem into several hierarchical *binary* sub-classifications. In each binary sub-classification, the HOCs are expanded into a higher dimensional space in which the two classes are linearly separable. It is shown that there is a significant improvement when using the proposed Hierarchical polynomial structure compared to the conventional polynomial classifier. Moreover, simulation results are shown for different block lengths (number of received symbols) and at different SNR values. The proposed system showed an overall improvement in the probability of correct classification that reaches 100% using only 512 received symbols at 20 dB compared to 98% and 98.33% when using more complicated systems like Genetic Programming with KNN classifier (GP-KNN) and Support Vector Machines (SVM) classifiers, respectively.

1. Introduction

Automatic Modulation Classification (AMC) is the process of identifying the modulation type of the transmitted signal from the received data samples automatically [1]; it is an intermediate step between signal detection and demodulation [2]. AMC has received a great deal of research and investigation in recent years because of its various applications in modern communication systems. For example, in cognitive radio (CR) systems, AMC is used to identify the types of signals in the spectrum. This information can be used to efficiently utilize the available spectrum and increase the data throughput [3]. In OFDM systems, AMC is used to identify the modulation type in each subcarrier. Accordingly, the appropriate demodulator is selected. On the other hand, the classical method of signaling the bit allocation table (BAT) to inform the receiver about the subcarrier modulation level degrades the system throughput significantly [4]. Moreover, AMC has been used in many military applications such as: spectrum surveillance, electronic warfare and threat analysis [5]. For example, AMC can be used to identify the modulation type of an intercepted enemies' signal so that the original transmitted message can be extracted [6].

AMC algorithms are divided into two main categories: Likelihood-Based (LB) schemes and Feature-Based (FB) schemes. The LB schemes deal with the automatic modulation classification problem as a multi-hypothesis problem in which a modulation type with the maximum likelihood among all the candidates will be assigned to the received signal [7]. Many types of LB algorithms were suggested in the literature, such as the Average Likelihood Ratio Test (ALRT) [8], Generalized Likelihood Ratio Test (GLRT) [8], Hybrid Likelihood Ratio Test (HLRT) [7] and Quasi-Hybrid Likelihood Ratio Test (QHLRT) [7]. Normally,

* Corresponding author.
E-mail addresses: b00050859@aus.edu.com (A. Abdelmutalab), kassaleh@aus.edu.com (K. Assaleh), mtarhuni@aus.edu.com (M. El-Tarhuni).

http://dx.doi.org/10.1016/j.phycom.2016.08.001
1874-4907/© 2016 Published by Elsevier B.V.

LB approaches are computationally complex and their performance degrades considerably in the existence of phase or frequency offset or impulsive noise [9].

In contrast, FB schemes are simpler to implement and can achieve a very close performance to LB schemes if the used features are chosen properly [10]. Different types of received signal features are suggested in the literature such as instantaneous amplitude, phase, and frequency [11], wavelet transform [12], Fourier transform [13], high order moments (HOMs), high order cumulants (HOCs) [5,14,15], higher-order cyclic cumulants [16], very high order statistics (VHOS) [17] and constellation diagram [18]. In general, the selection of the proper features depends mainly on the modulation types of interest. The selected features are used by a machine learning classifier to determine the modulation type of a received signal. Various types of classifiers are used in the field like Artificial Neural Networks (ANN) [19], Support Vector Machines (SVM) [20], Clustering Algorithms, K-Nearest Neighbors (KNN) [3], Polynomial Classifier (PC) [21], Threshold-Based Classifiers [14,22] and Naïve Bayes Classifier [23]. Furthermore, some work used optimization techniques such as Genetic Programming [3,24,25] and Particle Swarm Optimization (PSO) [20] in order to improve the classification features. However, classifiers differ in their complexity, accuracy and processing time. Another factor that limits a classifier's accuracy is the channel conditions through which the signal is transmitted. Many authors investigated AMC schemes in AWGN channels as in [23,24], whereas, other authors have considered more realistic channel models that took into consideration multipath fading [26].

Moreover, AMC can be implemented using one transmitter and one receiver, or one transmitter and multiple receivers. In the later method, the decision can be made by each receiver and then passed to a centralized system to vote, or all the received signals are directly processed by a centralized system [27].

It is noted that some modulation types are easier to classify such as BPSK and QPSK signals compared to modulation types with dense constellations like 64-QAM and 256-QAM. In both scenarios, the performance of the classifier is affected by the number of samples from which the features are extracted. Intuitively, the more the samples used in extracting the features the more accurate the system performs (for the case of AWGN channels). However, increasing the number of samples may reduce the system performance due to the unexpected phase drift, residual frequency offset, and timing error. This work is a continuation of our previous work proposed in [21] where only one multidimensional polynomial classifier was used to classify the different modulation types. It has been noticed that a severe drop in the probability of correct classification of the single polynomial classifier is observed when the number of considered modulation types increases. The probability of correct classification (PCC) is defined as the number of correctly classified signals divided by the total number of signals.

In this work, this problem is solved by using a tree structure of binary polynomial classifiers, where each classifier is trained to identify two classes at time. The proposed system has the following advantages:

- Provides higher probability of correct classification compared to most of the proposed systems in the literature.
- Has low computational complexity and can be easily implemented.
- Has an acceptable probability of correct classification for the case of slow Rician or Rayleigh fading.

The rest of the paper is organized as follows: Section 2 shows the signal model and introduces high order cumulants. Section 3 explains the system model, Section 4 shows the simulation results, whereas the conclusion is presented in Section 5.

2. Signal model and HOCs

In this work, the signals are assumed to be transmitted over a slow flat fading channel, which resembles realistic channel conditions especially when using OFDM systems to mitigate the effect of frequency selective fading channels. The baseband discrete-time received signal contaminated by Additive White Gaussian Noise (AWGN) in a flat fading environment can be expressed as:

$$y_n = h_n x_n + w_n, \quad n = 1, \ldots, N \tag{1}$$

where x_n is the discrete-time transmitted signal, w_n is the AWGN process with zero mean and two-sided power spectral density $\frac{N_0}{2}$, h_n is the complex-valued channel gain assumed to follow a Gaussian distribution, and N is the number of transmitted symbols per frame. The transmitted signal x_n is selected from L possible modulation types. In this work we consider BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM or 256-QAM modulation types (i.e. $L = 6$). Moreover, we assume that the noise variance is known or can be estimated at the receiver.

The multipath fading effect on the amplitude of the received signal is modeled using Rayleigh or Rician fading models. Furthermore, the relative motion between the transmitter and the receiver causes a shift in the frequency of the received signal f_d "Doppler Shift". The effect of different values of Doppler shift on the proposed automatic modulation classification scheme is investigated in this work.

The Higher order cumulants (HOCs) are estimated from the received signal and used as input features to the proposed classifier. Generally, HOCs are expressed as functions of the signals' High Order Moments (HOMs). For a complex-valued stationary random process $\mathbf{y}$, the pth order moment is defined as [22]

$$M_{pq} = E\left[\mathbf{y}^{p-q}(\mathbf{y}^*)^q\right] \tag{2}$$

where $\mathbf{y}^*$ is the complex conjugate of $\mathbf{y}$, and q is the power of the conjugate signal $\mathbf{y}^*$. Table 1 shows the relationship between the HOCs and HOMs [24].

Feature normalization is required in order for machine learning algorithms to work properly. For normalization, HOCs are rescaled as described in [28] with each cumulant raised to the power $\frac{2}{p}$, where p is the order of the cumulant, i.e. $\hat{C}_{42} = C_{42}^{\frac{1}{2}}$ and $\hat{C}_{63} = C_{63}^{\frac{1}{3}}$.

For more simplicity, the magnitudes of the cumulants are used instead of their complex values. This step has

Table 1
High order cumulants and high order moments proposed in [3].

HOCs		HOMs expression
Second order cumulants	C_{20}	M_{20}
	C_{21}	M_{21}
Fourth order cumulants	C_{40}	$M_{40} - 3M_{20}^2$
	C_{41}	$M_{40} - 3M_{20}M_{21}$
	C_{42}	$M_{42} - \|M_{20}\|^2 - 2M_{21}^2$
Sixth order cumulants	C_{60}	$M_{60} - 15M_{20}M_{40} + 30M_{20}^3$
	C_{61}	$M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30M_{20}^2M_{21}$
	C_{62}	$M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40} + 6M_{20}^2M_{22} + 24M_{21}^2M_{20}$
	C_{63}	$M_{63} - 9M_{21}M_{42} + 12M_{21}^3 - 3M_{20}M_{43} - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22}$

a great advantage in reducing the processing time in the training stage because the classifier weights are real-valued in this case instead of being complex. Another advantage of using the cumulants magnitude is reduced vulnerability to shifts in the constellation. As the phase shift does not affect the magnitude of the cumulants while it may affect their imaginary part [28]. Finally, we remark that some HOCs are more useful in separating a particular group of modulation types than the others, hence, it is desirable to decide on which HOCs to use automatically by the proposed scheme.

3. Proposed system

The objective of this work is to develop an AMC scheme that would allow the receiver to decide (classify) on which of the L possible modulation types is used by the transmitter to send a given block of data. In this work, AMC is treated as a pattern recognition problem, where features are extracted from the received signal and used in a classifier to decide upon the modulation level. The features used in this work are the HOCs of the received signal and the classifier used is a *hierarchical polynomial* (HP) classifier. It is noted that the performance of any pattern recognition scheme depends on the selected features applied to the classifier and the capability of the classifier itself. The performance of the classifier is measured by its probability of correct classification and computational complexity. In this section, we introduce the proposed HP classifier and discuss its training and testing stages.

3.1. Hierarchical polynomial classifier

The Weierstrass approximation theorem states that: "every real-valued continuous function on a finite closed interval [a, b] can be uniformly approximated by polynomials with real coefficients" and "every complex-valued continuous function on a finite closed interval [a, b] can be uniformly approximated by polynomials with complex coefficients" [29]. According to this theorem, a polynomial classifier can be used to approximate the nonlinear boundaries between different classes. A polynomial classifier is a machine learning algorithm that expands the original set of features in a given space to a higher dimensional space in which the different classes become linearly separated. Due to its simplicity and high accuracy, it has been used in variety of fields such as speech recognition [30], cognitive radio systems [31] and biomedical application [32].

Fig. 1 shows a binary classification example in the original features space d_1 and d_2. It can be noticed that the two classes are nonlinearly separable and can only be separated by a quadratic function. While, the new set of features x_1, x_2 and x_3 derived from an expansion of d_1 and d_2 can linearly separate the two classes in a higher dimensional space (3 dimensions in this case), where $x_1 = d_1^2$, $x_2 = d_2^2$ and $x_3 = \sqrt{2}d_1d_2$.

Similar to any supervised learning algorithm, a polynomial classifier has two main stages, as shown in Fig. 2. First, a *training* stage in which features from labeled training signals are used in calculating the classifier weights, and then a *testing* stage where unlabeled signals are applied to the classifier to identify the classes of these signals.

The proposed HP classifier is shown in Fig. 3, where the classification of the six modulation types is done in hierarchical binary-classifications stages. In each stage, modulations having features of similar values are clustered in one class and the rest are placed in a second class. Accordingly, the received signal is firstly classified to be PSK or QAM. In the next stage, if the signal is PSK, two new classes are introduced, BPSK as a class and QPSK and 8-PSK as another class. If the signal is classified as BPSK the classification procedure is completed, otherwise, another *binary* classification between QPSK and 8-PSK is introduced. The same principle applies for the QAM types as shown in Fig. 3. This process helps in optimizing the weights in each binary classification stage leading to an overall improvement in the classification accuracy. In contrast, the conventional polynomial classifier (non-hierarchical) used in [21] has less classification accuracy, especially for high values of L. To illustrate this, let us consider a case of six modulation types BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM, a conventional polynomial classifier will try to find the optimum set of weights for each modulation type, six weight vectors in this case. For example, the weight vector corresponding to BPSK modulation is calculated based on the assumption that BPSK is one class while QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM are another class. This is a valid assumption knowing that BPSK has a very different constellation compared to the others, accordingly, conventional polynomial classifier always gives the right decision when the transmitted signal is BPSK. In contrary, the weight vector corresponding to 64-QAM is calculated based on the assumption that 64-QAM is one class and BPSK, QPSK,

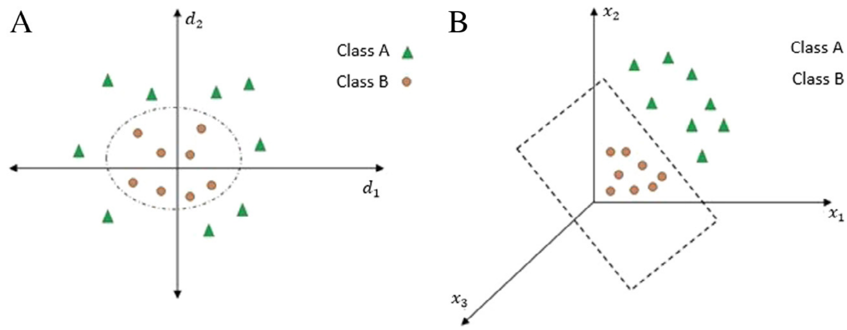
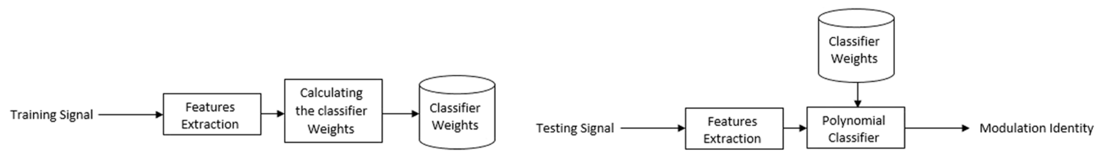


Fig. 1. Classification features in the original space and in the high dimensional space.



(a) Offline training. (b) Online testing.

Fig. 2. Training and testing stages.

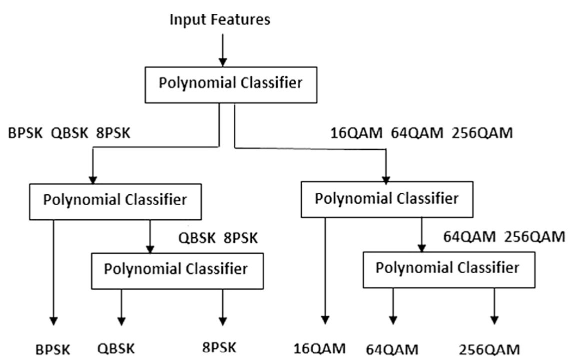


Fig. 3. Hierarchical polynomial classifier.

8-PSK, 16-QAM and 256-QAM are another class. The problem arises in this case, because the classifier assumes that very different modulation types like BPSK and 256-QAM belong to the same class while 64-QAM which is very similar to 256-QAM is a totally different class. This is why conventional polynomial classifier tends to miss-classify 64-QAM as 256-QAM. The hierarchical polynomial classifier solved this problem by using a single binary classification at a time, two set of weights are calculated to classify 64-QAM and 256-QAM, and there is no confusion introduced by using the other modulation types in the training.

3.2. Training stage

Second, fourth and sixth order cumulants are used as features from each received signal with known modulation type. The obtained features are expanded into a higher dimensional space using a polynomial classifier in order to produce more features and allow for easier separation of the classes. The order of the polynomial classifier determines the dimensionality of the space. Although higher order polynomial classifiers could be used, but for implementation simplicity, a second order polynomial classifier is normally implemented, where the new set of features is the original features plus their products and squared values. For instance, for an input feature vector $\mathbf{d}$ composed of normalized HOCs defined as $\mathbf{d} = [d_1, d_2, \ldots, d_M]$, the expanded feature vector $\mathbf{p}$ is expressed as [21]:

$$\mathbf{p} = \begin{bmatrix} 1,\ d_1, \ldots, d_M,\ d_1 \times d_2,\ \ldots,\ d_1 \times d_M, \\ d_2 \times d_3,\ \ldots,\ d_2 \times d_M,\ \ldots, d_{M-1} \times d_M, \\ d_1^2,\ d_2^2,\ \ldots,\ d_M^2 \end{bmatrix}_{1\times R} \tag{3}$$

where M is the number of HOCs used as input features, and R is the dimension of the expanded feature vector.

The proposed scheme is capable of classifying any number of modulation schemes using the appropriate hierarchical approach with two classes at time. To explain how a polynomial classifier works, let us consider a case of two modulation types, for example BPSK and QPSK. In order to train the polynomial classifier, K training signals are transmitted using BPSK and another K training signals using QPSK modulations. Second, fourth and sixth order cumulants are estimated from each training signal and the corresponding $\mathbf{p}$ vector is calculated using (3). After that, a new matrix $\mathbf{v}_l$ that has all the expanded feature vectors for modulation type l is defined as:

$$\mathbf{v}_l = \begin{bmatrix} \mathbf{p}_1^{(l)} \\ \mathbf{p}_2^{(l)} \\ \vdots \\ \mathbf{p}_K^{(l)} \end{bmatrix}_{K\times R} \qquad l = 1, 2. \tag{4}$$

By rearranging both of $\mathbf{v}_1$ for BPSK modulation and $\mathbf{v}_2$ for QPSK in a new matrix $\mathbf{V}$, we get:

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix}_{2K\times R}. \tag{5}$$

Next step is to calculate the classifier weights, $\mathbf{w}_1^o$ for BPSK and $\mathbf{w}_2^o$ for QPSK. In order to do so, the weights are selected to minimize the mean square error as:

$$\mathbf{w}_l^o = \arg\min_{\mathbf{w}_l} \|\mathbf{V}\mathbf{w}_l - \mathbf{t}_l\|^2 \tag{6}$$

where $\mathbf{w}_l$ is an $R \times 1$ vector and $\mathbf{t}_l$ is the target matrix defined as:

$$\mathbf{t}_1 = \begin{bmatrix}\mathbf{1}\\ \mathbf{0}\end{bmatrix}_{2K\times 1}, \quad \text{when calculating } \mathbf{w}_1^o \tag{7}$$

$$\mathbf{t}_2 = \begin{bmatrix}\mathbf{0}\\ \mathbf{1}\end{bmatrix}_{2K\times 1}, \quad \text{when calculating } \mathbf{w}_2^o \tag{8}$$

and

$$\mathbf{0} = [0, 0, \dots, 0]^{\mathbf{T}}_{1\times K} \tag{9}$$

$$\mathbf{1} = [1, 1, \dots, 1]^{\mathbf{T}}_{1\times K}. \tag{10}$$

Eq. (6) can be simply written as:

$$\mathbf{V}\mathbf{w}_l^o = \mathbf{t}_l. \tag{11}$$

Then:

$$\mathbf{w}_l^o = \mathbf{V}^{\dagger}\mathbf{t}_l \tag{12}$$

where $\mathbf{V}^{\dagger}$ is the pseudo inverse of matrix $\mathbf{V}$ [33]. Finally, after finding $\mathbf{w}_1^o$ and $\mathbf{w}_2^o$, the classifier is ready for the testing stage.

3.3. Testing stage

In the testing stage, the objective is to find the identity of an unlabeled modulated signal (either BPSK or QPSK for the example under consideration). To classify the modulation type, the HOCs feature vector $\mathbf{d}$ is first extracted from the received signal and the expanded vector $\mathbf{p}$ is calculated using the second order expansion in (3). Then, vector $\mathbf{p}$ is multiplied by classifier weights $\mathbf{w}_1^o$ and $\mathbf{w}_2^o$ obtained during the training stage to give the scores s_1 and s_2, respectively, to decide on the modulation type. Ideally, the weights are optimized during the training stage to give $s_1 = 1$ and $s_2 = 0$ if the modulation is type one (BPSK), and to give $s_1 = 0$ and $s_2 = 1$ if the modulation is type two (QPSK). Since the received symbols are noisy, the decision is made based on the maximum values of s_1 and s_2, meaning that if $s_1 > s_2$ then the modulation is BPSK, and if $s_2 > s_1$ the modulation is QPSK. Hence,

$$\textit{Class identity } l = \arg\max_l \{s_l\}. \tag{13}$$

The previous example explained the modulation classification process for only two modulation types, BPSK and QPSK. In this work, we consider a more general classification problem among all of BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM. Using the same concept introduced before, we divide the modulation classification process into binary sub-classification stages. In the first stage PSK signals are treated as one class and QAM signals as another class. In other words, we define a problem of two modulation groups, PSK modulation group that has BPSK, QPSK and 8-PSK signals, and QAM modulation group that has 16-QAM, 64-QAM and 256-QAM. K training signals form BPSK, QPSK and 8-PSK are used to calculated matrix $\mathbf{v}_1$, and K training signals form 16-QAM, 64-QAM and 256-QAM are used to calculated matrix $\mathbf{v}_2$. Following the same procedure, $\mathbf{w}_1^o$ and $\mathbf{w}_2^o$ are calculated in the training stage. In the testing stage if $s_1 > s_2$ the received signal is PSK and if $s_2 > s_1$ the received signal is QAM. If the signal is decided to be PSK, another binary modulation classification problem is considered. This time, another polynomial classifier is used where matrix $\mathbf{v}_1$ has the expanded feature vector of BPSK modulation, and matrix $\mathbf{v}_2$ has the expanded feature vector of both QPSK and 8-PSK. Accordingly, in the testing stage, if $s_1 > s_2$ the received signal is BPSK and if $s_2 > s_1$ the received signal is either QPSK or 8-PSK, and another polynomial classifier is used to determine whether the modulation type is QPSK or 8-PSK using the same illustrated concept. The same idea is used if the signal is decided to be QAM, as shown in Fig. 3.

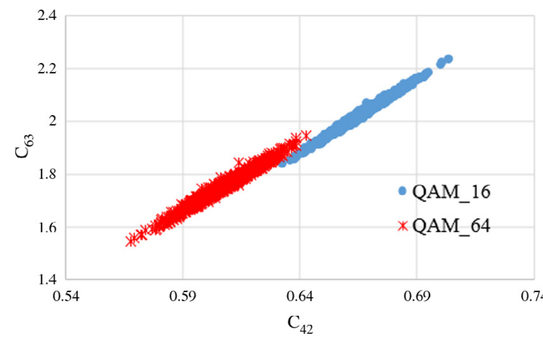


Fig. 4. Features in the original dimensional space.

4. Simulation results

In this section, the performance of the proposed AMC scheme is examined under different channel conditions and is compared to other methods from the literature. First, the advantage of the proposed classifier over the traditional threshold-based method is investigated. A simulation involving the generation of 1000 different realizations of 16-QAM and 64-QAM signals is conducted at 20 dB signal-to-noise (SNR), with each signal represents a block of $N = 2000$ symbols. The conventional method for classifying 16-QAM and 64-QAM signals uses the value of fourth order cumulant C_{42} or the value of the sixth order cumulant C_{63} as defined in [34]. Using a threshold to decide on the modulation type is not the optimal solution. Fig. 4 shows that, there is no optimum threshold value for C_{42} or C_{63} that can accurately separate the two modulation types. On the other hand, Fig. 5 shows the two output scores when using the proposed polynomial classifier with the same input features C_{42} and C_{63}. In this scenario, when 16-QAM is transmitted (the first 1000 signals), then the first score s_1 is always greater than the second score s_2. The opposite happens when the transmitted signal is 64-QAM (signals numbered 1001–2000), where the second score s_2 is always greater than the first score s_1. Using the values of s_1 and s_2, a probability of correct classification close to 100% is obtained.

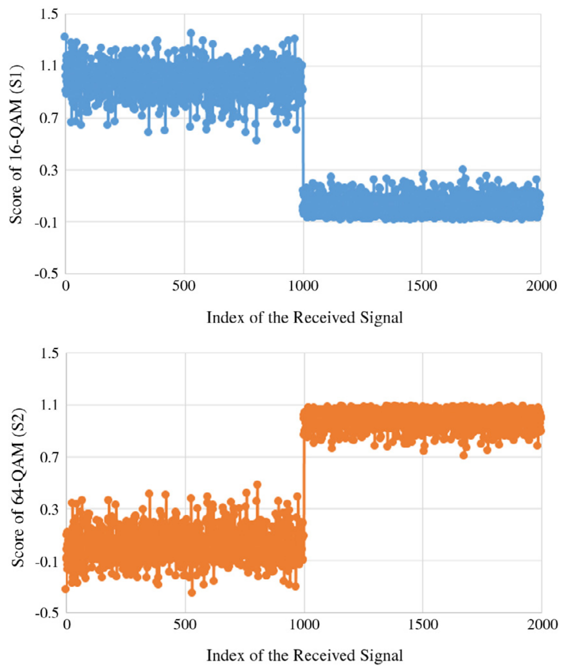


Fig. 5. The two scores (super features) of the polynomial classifier.

4.1. Hierarchical polynomial classifier

In this section and the next section, the expanded feature vector **p** is calculated from the nine HOCs in Table 1 and used to distinguish between the different modulation types. In Fig. 6, a classification problem among BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM is considered under AWGN channel conditions. First, we have compared the performance of the proposed HP classifier with the *conventional polynomial* (CP), i.e. nonhierarchical, classifier, under the assumption that no SNR information is available at the receiver side. During the training stage, signals with different known SNR values are used to calculate an average classifier weights. The average weights are then used during the testing stage to classify the modulation type of unknown signals regardless of the received signal SNR. The results show that the HP classifier provides better correct classification of the modulation type than the CP scheme, especially at higher values of SNR. However, due to the lack of SNR information, the performance of both schemes reaches an upper limit indicating the non-optimality of selecting the classifier weights. Then, we investigated the case when the SNR is known at the receiver. Accordingly, classifier weights are calculated for each SNR value in the training stage. Then, based on the estimated SNR of the received signal, the corresponding classifier weights are used in the testing stage. The results presented in Fig. 6 show that when perfect SNR information is available at the receiver then significant improvement in the classification accuracy can be achieved; reaching to about 100% accuracy for SNR above 12 dB. The advantage of using the proposed HP over the CP in the two scenarios is clearly demonstrated.

The effect of using different number of symbols per data block on the probability of correct classification of HP classifier is shown in Fig. 7. It is shown that using a larger

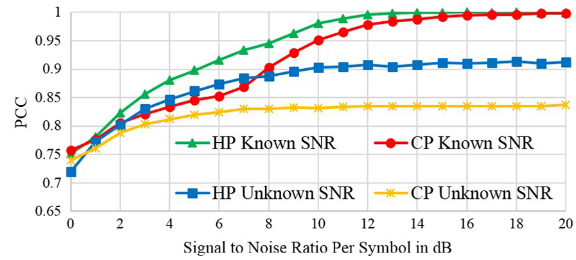


Fig. 6. Probability of correct classification in AWGN only for CP and HP classifiers.

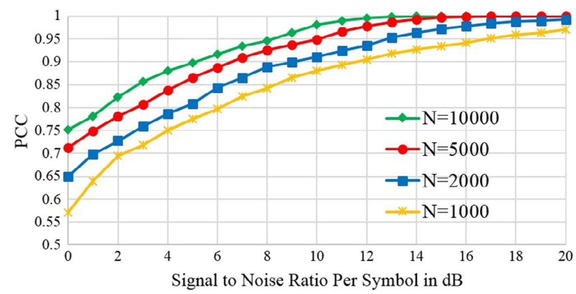


Fig. 7. Probability of correct classification in AWGN only for different number of symbols.

number of received symbols to extract the classification features improves the probability of correct classification. For example, using $N = 5000$ symbols leads to a 4 dB improvement in SNR compared to $N = 1000$ symbols. From Figs. 6 and 7, it is clear that the probability of correct classification is a function of SNR of the channel and signal's block length, this is due to the variation in HOCs values based on these parameters. While HOCs have average constant values (usually considered as a reference to identify the modulation type), the variation from these average values increases with low SNR and small block lengths resulting in signals misclassification [35].

Table 2 shows the confusion matrix for the six modulation types (BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 245-QAM) when $N = 10\,000$ and at SNR of 10 dB. Each row represents the actual modulation type (2000 signals are transmitted per each modulation type), and the column represents the predicted modulation identity by the classifier. For example, when sending 2000 signals of modulation type 16-QAM, the system was able to recognize it as 16-QAM all the time. Whereas, when sending 2000 signals of modulation type 64-QAM, The system classified 1887 of them as 64-QAM and misclassified 113 signals as 256-QAM. Table 3 shows the confusion matrix for $N = 5000$ and at SNR of 5 dB. It is clear that the number of misclassified signals in this case is higher compared to Table 2. This is expected, since the number of transmitted symbols per signal is less than the case in Table 2 and the SNR is lower. Moreover, Most of the modulation misclassification is between 64-QAM and 256-QAM, since they have very similar constellations especially at low SNR. However, for the sake of simplicity and to be able to compare our work to the published work in the literature, we calculate the average probability of correct classification for the six modulation types, which equals to 98.04% in Table 2 and 86.52% in Table 3.

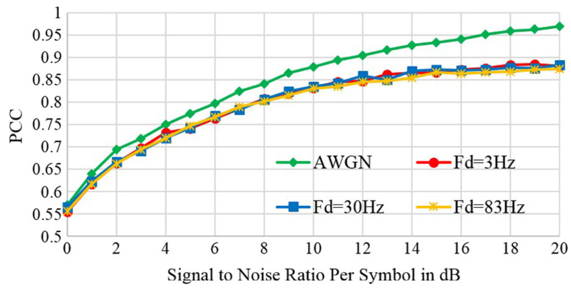


Fig. 8. Probability of correct classification in slow Rician fading channels.

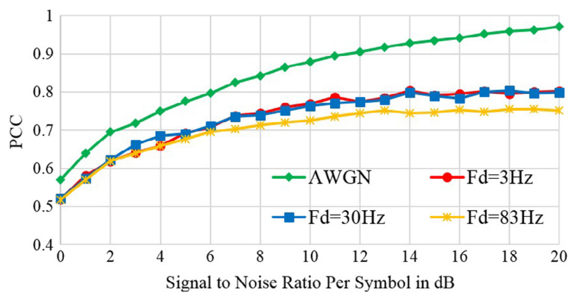


Fig. 9. Probability of correct classification in slow Rayleigh fading channels.

The performance of the proposed scheme is investigated in slow flat fading channel conditions with magnitude follows Rayleigh (no line of sight) or Rician fading (with line of sight component) whereas the phase is uniformly distributed and is assumed to be constant during the symbol period. Without loss of generality, the Rician factor, which represents the ratio between the power of the direct path and the power of the reflected paths, is assumed to be 6 dB. The parameters used are the same as in [28] with a symbol rate of 1×10^6 symbols per second, average path gains $[0, \quad -10]$ dB, path delays $\left[0, \quad 1 \times 10^{-7}\right]$ s, and maximum Doppler shift of 3 Hz, 30 Hz, or 83 Hz. Figs. 8 and 9 show the probability of correct classification for $N = 5000$ with different values of maximum Doppler shifts in Rician and Rayleigh fading, respectively. The maximum Doppler shift and the SNR of the channel are assumed to be known at the receiver side. Accordingly these values are used to train the classifier. It is clear that the probability of correct classification is higher for Rician fading channels compared to Rayleigh fading channels. This is because of the existence of the strong signal due to the line of sight component. It is noted that a higher Doppler shift results in lower probability of correct classification. Finally, Fig. 10 shows The classification performance as a function of the number of received symbols and Doppler shift for flat Rayleigh fading channels.

4.2. Comparison to related work

Table 4 compares the probability of correct classification of the Naïve Bayes classifier, Support Vector Machines (SVM) classifier, Genetic programming with KNN classifier and the proposed Hierarchical Polynomial (HP) classifier. To ensure fair comparison, the proposed classifier is simulated using the same parameters used by the other authors.

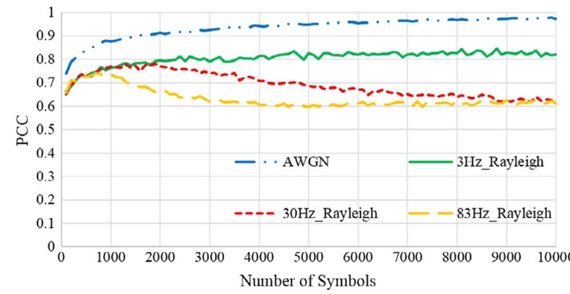


Fig. 10. Probability of correct classification versus the number of symbols for different values of Doppler shift.

In particular, we used four modulation types (BPSK, QPSK, 16-QAM and 64-QAM) with 512, 1024 and 2048 symbols per signal. The proposed HP classifier achieved a probability of correct classification of 100% at 20 dB for the three given scenarios $N = 512$, 1024 and 2048 symbols. For $N = 512$ at 10 dB HP classifier managed to achieve a 96.49% accuracy, whereas, GP-KNN, SVM and Naïve Bayes classifiers reported accuracy of 94%, 91.23% and 90.17%, respectively. It is clear that the proposed HP classifier outperformed the three other classifiers in terms of probability of correct classification. However, the probability of correct classification is not the only factor that favors one classifier over another. A crucial factor is the system complexity and the order of calculations involved in making the classifier decision. Since, modulation classification is a real-time identification problem, systems with simple structure and low complexity are preferred.

The complexity of a classifier can be analyzed based on the complexity of the training stage (offline) and the complexity of the testing stage (online). In the training stage, GP-KNN and SVM use iterative approaches and consume significant processing time to provide the final training model, especially for the case of GP-KNN where different function pools are used to calculate the final training model. However, for our proposed system and Naïve Bayes classifier, the training process is straight forward without any iterations involved. In general, the complexity of the training stage does not usually matter if the calculations in the testing stage are simple, because the training is performed offline.

In the testing stage, nine features are applied to the proposed HP classifier, and the classifier expands them forming a new feature vector with 55 features. It is noted that the only complexity in this method is calculating the expanded feature vector and then multiplying it by the classifier weights. The complexity of the proposed HP classifier can be estimated as $O(2f)$ where f is the dimension of the expanded feature vector, and the factor 2 is for the two binary sub-classification required for each decision. This complexity is constant even if the number of the training samples is high. On the other hand, for the case of the GP-KNN system, the complexity is much higher, since it contains the complexity of finding the final super feature (the calculations to find this super feature depends on the pool function) used like (*plus, minus, times, reciprocal, negator, abs, sqrt, sin, cos, tan, asin, acos, tanh*) [24]. Furthermore, in GP-KNN classifier, the distance

Table 2
Confusion matrix for $N = 10\,000$ at SNR = 10 dB (98.04%).

	BPSK	QPSK	8-PSK	16-QAM	64-QAM	256-QAM
BPSK	2000	0	0	0	0	0
QPSK	0	2000	0	0	0	0
8-PSK	0	0	2000	0	0	0
16-QAM	0	0	0	2000	0	0
64-QAM	0	0	0	0	1887	113
64-QAM	0	0	0	0	122	1878

Table 3
Confusion matrix for $N = 5000$ at SNR = 5 dB (86.52%).

	BPSK	QPSK	8-PSK	16-QAM	64-QAM	256-QAM
BPSK	2000	0	0	0	0	0
QPSK	0	2000	0	0	0	0
8-PSK	0	0	2000	0	0	0
16-QAM	0	0	0	1811	189	0
64-QAM	0	0	0	29	1276	695
64-QAM	0	0	0	3	701	1296

Table 4
Comparison to other systems in the literature.

N	SNR (dB)	Probability of correct classification (%)			
		Naïve [23]	SVM [23]	GP-KNN [3]	HP (proposed method)
	0	63.91	64.53	65	65.78
512	10	90.17	91.23	94	96.49
	20	94.66	98.33	98	100
	0	69.68	70.3	70	71.99
1024	10	94.4	94.81	98	99.10
	20	98.28	98.89	100	100
	0	76.75	75.73	75	77.60
2048	10	97.89	97.92	100	99.96
	20	99.68	99.78	100	100

is calculated between the calculated super feature and all the training samples; for example the number of the reference points in [36] is 400 points, accordingly, after finding the super feature the distances between this super feature and the 400 reference points are calculated. Hence, the overall complexity of GP-KNN is $O(N)$ where $N = 400$ plus the complexity of generating the super feature.

For the SVM and Naïve Bayes classifiers in the testing stage the complexity is less than GP-KNN classifier, but their overall classification accuracy is relatively lower. Finally, it is rational that the more complex computations the classifier performs the higher probability of correct classification it provides. Yet in our proposed system, we managed to achieve high classification accuracy and maintain simple classifier structure.

5. Conclusion

In this paper, a hierarchical polynomial classifier is used for automatic modulation classification of M-PSK and M-QAM signals. Higher order cumulants features are extracted from the received signal and applied to the proposed classifier. The proposed scheme expands the original feature vector into a higher dimensional space in which the classes are easily classified. It has been shown that using the channel information at the receiver side to train the classifier improves the probability of correct classification significantly. Moreover, using higher number of received symbols to extract the classification features results in a higher probability of correct classification. For channels with slow flat fading, the proposed system showed no degradation due to the phase shift in the constellation. Then, a relationship between the number of the received symbols used to extract the classification features and the probability of correct classification is investigated for different scenarios of Doppler shifts. Finally, the advantage of the proposed system is investigated in terms of accuracy and calculations complexity compared to other work in the literature.

References

[1] V. Orlic, M. Dukic, Algorithm for automatic modulation classification in multipath channel based on sixth-order cumulants, in: 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services, TELSIKS'09, 2009, pp. 423–426.

[2] C. Yin, B. Li, Y. Li, B. Lan, Modulation classification of MQAM signals based on density spectrum of the constellations, in: 2nd International Conference on Future Computer and Communication, ICFCC, Vol. 3, 2010, pp. V3–57–V3–61.

[3] M. Aslam, Z. Zhu, A. Nandi, Automatic modulation classification using combination of Genetic Programming and KNN, IEEE Trans. Wireless Commun. 11 (8) (2012) 2742–2750.

[4] L. Häring, Y. Chen, A. Czylwik, Automatic modulation classification methods for wireless OFDM systems in TDD mode, IEEE Trans. Commun. 58 (9) (2010) 2480–2485.

[5] L. Liu, J. Xu, A novel modulation classification method based on high order cumulants, in: International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM, 2006, pp. 1–5.

[6] B. Ramkumar, Automatic modulation classification and blind equalization for cognitive radios (Ph.D. thesis), Virginia Polytechnic Institute and State University, Virginia.
[7] F. Hameed, O. Dobre, D. Popescu, On the likelihood-based approach to modulation classification, IEEE Trans. Wireless Commun. 8 (12) (2009) 5884–5892.
[8] O. Dobre, A. Abdi, Y. Bar-Ness, W. Su, Survey of automatic modulation classification techniques: classical approaches and new trends, IET Commun. 1 (2) (2007) 137–156.
[9] F. Wang, X. Wang, Fast and robust modulation classification via Kolmogorov-Smirnov test, IEEE Trans. Commun. 58 (8) (2010) 2324–2332.
[10] A. Hazza, M. Shoaib, S. Alshebeili, A. Fahad, An overview of feature-based methods for digital modulation classification, in: 1st International Conference on Communications, Signal Processing, and their Applications, ICCSPA, 2013, pp. 1–6.
[11] K. Assaleh, K. Farrell, R. Mammone, A new method of modulation classification for digitally modulated signals, in: IEEE Military Communications Conference, MILCOM'92, Vol. 2, 1992, pp. 712–716.
[12] W. Dan, G. Xuemai, G. Qing, A new scheme of automatic modulation classification using wavelet and wsvm, in: 2nd International Conference on Mobile Technology, Applications and Systems, 2005, p. 5.
[13] F. Xie, C. Li, G. Wan, An efficient and simple method of MPSK modulation classification, in: 4th International Conference on Wireless Communications, Networking and Mobile Computing WiCOM'08, 2008, pp. 1–3.
[14] A. Swami, B. Sadler, Hierarchical digital modulation classification using cumulants, IEEE Trans. Commun. 48 (3) (2000) 416–429.
[15] Y. Han, G. Wei, C. Song, L. Lai, Hierarchical digital modulation recognition based on higher-order cumulants, in: 2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control, IMCCC, 2012, pp. 1645–1648. http://dx.doi.org/10.1109/IMCCC.2012.398.
[16] O. Dobre, Y. Bar-Ness, W. Su, Higher-order cyclic cumulants for high order modulation classification, in: Military Communications Conference, MILCOM, Vol. 1, 2003, pp. 112–117.
[17] W. Su, Feature space analysis of modulation classification using very high-order statistics, IEEE Commun. Lett. 17 (9) (2013) 1688–1691.
[18] C. Zhendong, J. Weining, X. Changbo, L. Min, Modulation recognition based on constellation diagram for M-QAM signals, in: 11th IEEE International Conference on Electronic Measurement Instruments, ICEMI, Vol. 1, 2013, pp. 70–74.
[19] J. Popoola, R. Van Olst, Automatic classification of combined analog and digital modulation schemes using feedforward neural network, in: AFRICON, 2011, pp. 1–6.
[20] M. Valipour, M. Homayounpour, M. Mehralian, Automatic digital modulation recognition in presence of noise using SVM and PSO, in: Sixth International Symposium on Telecommunications, IST, 2012, pp. 378–382.
[21] A. Abdelmutalab, K. Assaleh, M. El-Tarhuni, Automatic modulation classification using polynomial classifiers, in: 25th IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC, 2014, pp. 785–789.
[22] M. Mirarab, M. Sobhani, Robust modulation classification for PSK /QAM/ASK using higher-order cumulants, in: 6th International Conference on Information, Communications Signal Processing, 2007, pp. 1–4.
[23] M. Wong, S.K. Ting, A. Nandi, Naïve Bayes classification of adaptive broadband wireless modulation schemes with higher order cumulants, in: 2nd International Conference on Signal Processing and Communication Systems, ICSPCS, 2008, pp. 1–5.
[24] M. Aslam, Z. Zhu, A. Nandi, Automatic digital modulation classification using Genetic Programming with K-Nearest Neighbor, in: Military Communications Conference, MILCOM, 2010, pp. 1731–1736.
[25] N. Ahmadi, R. Berangi, Modulation classification of QAM and PSK from their constellation using genetic algorithm and hierarchical clustering, in: 3rd International Conference on Information and Communication Technologies: From Theory to Applications, ICTTA, 2008, pp. 1–5.
[26] H.-C. Wu, M. Saquib, Z. Yun, Novel automatic modulation classification using cumulant features for communications via multipath channels, IEEE Trans. Wireless Commun. 7 (8) (2008) 3098–3105.
[27] W. Su, Modulation classification of single-input multiple-output signals using asynchronous sensors, IEEE Sens. J. 15 (1) (2015) 346–357.
[28] N.P. Geisinger, Classification of digital modulation schemes using linear and nonlinear classifiers (Master's thesis), Naval Postgraduate School, California, 2010.
[29] P. Borwein, T. Erdelyi, Polynomials and Polynomial Inequalities, Springer, 1995.
[30] W. Campbell, K. Assaleh, C. Broun, Speaker recognition with polynomial classifiers, IEEE Trans. Speech Audio Process. 10 (4) (2002) 205–212.
[31] Y. Hassan, M. El-Tarhuni, K. Assaleh, Comparison of linear and polynomial classifiers for co-operative cognitive radio networks, in: 21st IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC, 2010, pp. 797–802.
[32] K. Assaleh, H. Al-Nashash, A novel technique for the extraction of fetal ECG using polynomial networks, IEEE Trans. Biomed. Eng. 52 (6) (2005) 1148–1152.
[33] T. Shanableh, K. Assaleh, Feature modeling using polynomial classifiers and stepwise regression, Neurocomputing 73 (10–12) (2010) 1752–1759.
[34] V. Orlic, M. Dukic, Automatic modulation classification: Sixth-order cumulant features as a solution for real-world challenges, in: 20th Telecommunications Forum, TELFOR, 2012, pp. 392–399.
[35] A. Abdelmutalab, Learning-based automatic modulation classification (Master's thesis), American University of Sharjah, 2015.
[36] M. Aslam, Pattern recognition using genetic programming for classification of diabetes and modulation data (Ph.D. thesis), the University of Liverpool, Liverpool, 2013.

Ameen Abdelmutalab received the B.Sc. degree from the University of Khartoum, Sudan in 2011 and the M.Sc. degree from the American University of Sharjah, UAE in 2015. He is currently a Ph.D. candidate at Concordia University, Montreal, Canada. After completing his B.Sc., he worked as a teaching assistant at the University of Khartoum and the University of Science and Technology in Sudan. He also worked as a graduate teaching assistant during his studies at the American University of Sharjah. His research interests include wireless communications, cognitive radio, machine-learning applications and image and video processing.

Khaled Assaleh is currently a Professor of electrical engineering at the American University of Sharjah, UAE where he is also the interim vice provost for research and graduate studies. He received his Ph.D. in electrical engineering from Rutgers, The State University of New Jersey in 1993. Upon completing his Ph.D. he worked as a Research Professor at the CAIP Center of Rutgers University for one year. He then had an 8-year career in industry with Motorola, Inc. in Phoenix, Arizona and with Rockwell Semiconductor Systems (Conexant/Skyworks) in Newport Beach, California. Assaleh holds 12 US patents and has published over 100 articles in signal processing and pattern recognition and their applications. His research interests include bio-signal processing, biometrics, speech processing, and pattern recognition. Dr. Assaleh has served in the organization committees of several international conferences in different capacities. He has also served as a guest editor for several journal special issues.

Mohamed El-Tarhuni received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Benghazi, Libya, in 1986 and 1990, respectively, and the Ph.D. degree in electrical engineering from Carleton University, Ottawa, Canada, in 1997. From 1987 to 1993, he was with the General Electric Company of Libya, Benghazi, Libya, as a Telecommunications Engineer and Manager of the Communications Department. From 1997 to 2000, he was with Nortel Networks, Ottawa, ON, Canada, as a member of Scientific Staff, working on third-generation wireless communication systems using CDMA technology. He has been with the American University of Sharjah (AUS), Sharjah, United Arab Emirates (UAE), since 2000, where he is currently a Professor and Associate Dean for the College of Engineering. From 2001 to 2006, he was the Director for the Cisco Regional Networking Academy, AUS. He was the Head of Electrical Engineering Department from 2006 to 2014. His current research interests include wireless and mobile radio systems, cognitive radio, and physical layer issues. Dr. El-Tarhuni has served on the organizing and technical committees of many international conferences. He served as the IEEE Industry Relations Officer for the IEEE UAE Section from 2002 to 2006. He is a senior member of the IEEE.